

使用 lung-cancer-study 資料檔。

1. 計算肺癌病人中接受化療 (CHEMO=1) 的存活曲線的 Kaplan-Meier 估計及圖。
2. 接第 1 題，使用 Log-log 轉換及 log 轉換分別計算存活曲線估計中每點的 95% 信賴區間，並比較差異？
3. 接第 1 題，計算存活曲線的中位數及  $\frac{1}{4}$  及  $\frac{3}{4}$  分位數？
4. 接續問題 1~3，計算肺癌病人中接受放療 (RT=1) 的存活表現，並檢定 " 僅 " 接受放療和僅接受化療病人間存活的差異。
5. 將復發(relapse)到死亡時間當成 " 存活時間 "，使用 COX 模型(I).做單一變數的分析，探討年齡( "年齡減年齡中位數" 當做年齡變數)，性別、抽煙、化療、T-stage 是否為風險因子？風險比各為多少？95%信賴區間?(II).使用上述因子作多變量分析，並做出結論。(III).以性別做分層分析並給出結論。

1. 計算肺癌病人中接受化療( CHEMO=1 )的存活曲線的Kaplan-Meier估計及圖。
2. 接第1題，使用Log-log轉換及log轉換分別計算存活曲線估計中每點的95%信賴區間，並比較差異？
3. 接第1題，計算存活曲線的中位數及 $\frac{1}{4}$ 及 $\frac{3}{4}$ 分位數？

答：

此處將問題1~3合併分析再個別說明

(1) 此處僅使用接受化療的人(CHEMO=1, YES)作為分析資料，使用資料處理中的資料篩選得到資料共94筆，並另存新檔lung\_cancer\_study\_chemo\_1，篩選過程此處略過。

(2) 計算Kaplan-Meier估計可使用R-web中的存活分析，在存活分析中需有時間與事件兩個變數，此處的時間變數為SURVIVAL\_MONTH，而事件變數是vital\_status。

R-web 分析步驟：

分析方法→存活分析→Kaplan-Meier 存活函數估計

→步驟一：資料匯入(選取個人資料檔 lung\_cancer\_study\_chemo\_1)

→步驟二：參數設定(時間變數：SURVIVAL\_MONTH、事件變數：vital\_status)

→進階選項(選擇信賴區間方法：Log-log 與 Log)

→開始分析

## 分析結果

- 分析方法：Kaplan-Meier 存活函數估計
- 資料名稱：lung\_cancer\_study\_chemo\_1
- 時間變數：SURVIVAL\_MONTHS
- 事件變數：vital\_status (設限指標：0)
- 顯著水準：0.05
- 信賴區間：Log-log
- 計算時間：0.276 秒

表 1 K-M 存活率估計及 Log-log(Log) 轉換的 95%信賴區間(CHEMO=1)

時間 time	涉險人數 no. at risk	事件人數 no. of event	K-M 存活 率估計 K-M survival	標準差 std. err	95 % 信賴區間		95 % 信賴區間	
					Log		Log-log	
					下界 lower	上界 upper	下界 lower	上界 upper
2.27	92	1	0.9891	0.0108	0.9682	1	0.9253	0.9985
6	91	1	0.9783	0.0152	0.9489	1	0.9159	0.9945
7.69	90	1	0.9674	0.0185	0.9318	1	0.9023	0.9894
8	89	1	0.9565	0.0213	0.9157	0.9991	0.8883	0.9835
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
96.1	9	1	0.2139	0.0531	0.1315	0.348	0.1207	0.3247
114	4	1	0.1604	0.0611	0.0761	0.3383	0.0638	0.2961
115.8	3	1	0.1069	0.0597	0.0358	0.3194	0.0261	0.254
132.5	1	1	0	NaN	NA	NA	NA	NA

表 2 百分位數估計值摘要(CHEMO=1)

參數 parameters	估計 estimation
25 百分位數	95
50 百分位數	51
75 百分位數	20.6

圖 1 存活函數估計圖及 Log-log 轉換計算的%95 信賴區間(CHEMO=1)

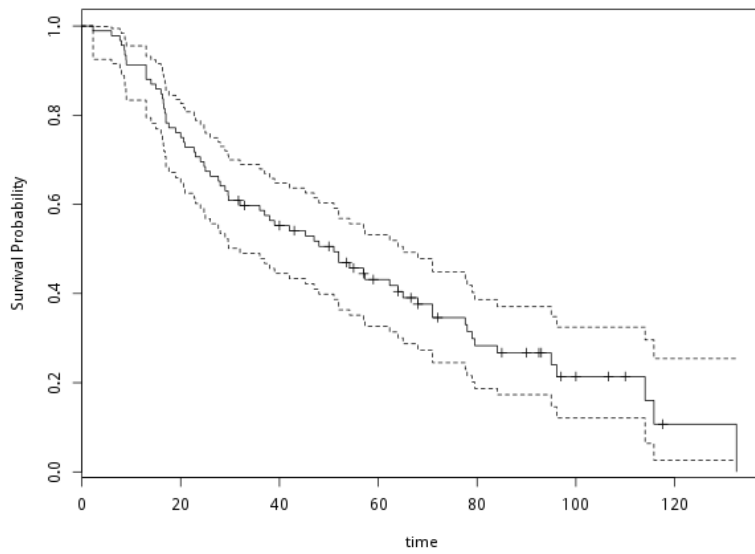
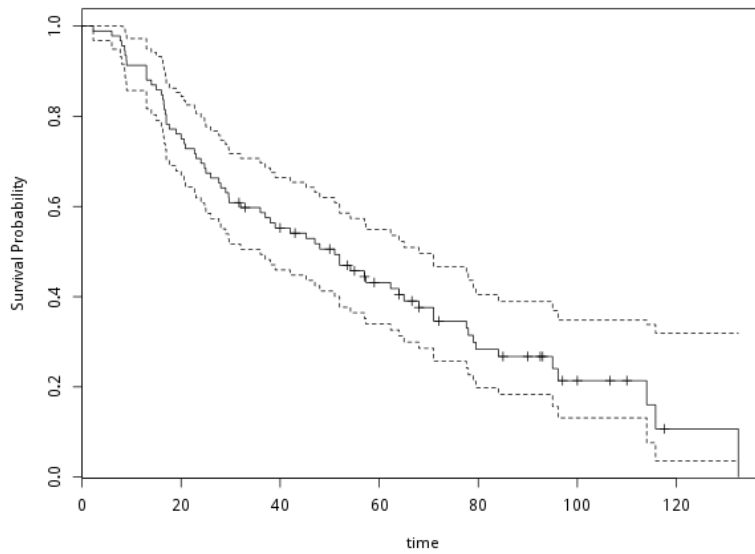


圖 2 存活函數估計圖及 Log 轉換計算的%95 信賴區間(CHEMO=1)



小結：

(1) 肺癌病人中接受化療 ( CHEMO=1 ) 的存活曲線的 Kaplan-Meier 估計及圖分列於表 1 及圖 1、2。

(2) 使用 Log-log 轉換及 log 轉換分別計算存活曲線估計中每點的 95%信賴區間，結果列於表 1 中，可發現 Log 轉換的信賴區間較 Log-log 轉換的信賴區間為高，但兩者的寬度並沒有絕對的大小差異。

(3) 存活曲線的中位數及 $\frac{1}{4}$ 及 $\frac{3}{4}$ 分位數列於表 2 中。

4. 接續問題 1~3，計算肺癌病人中接受放療 ( RT=1 ) 的存活表現，並檢定 " 僅 " 接受放療和僅接受化療病人間存活的差異。

答：

(1)此處需使用肺癌病人中接受放療( RT=1 )的資料再重新接續 1~3 題分析乙次。處理方法同前，將問題 1~3 合併分析再個別說明

使用接受化療的人(RT=YES，rt=1)作為分析資料，使用資料處理中的資料篩選得到資料共 73 筆，並另存新檔 lung\_cancer\_study\_RT\_1，篩選過程此處略過。

計算 Kaplan-Meier 估計可使用 R-web 中的存活分析，在存活分析中需有時間與事件兩個變數，此處的時間變數為 SURVIVAL\_MONTH，而事件變數是 vital\_status。

R-web 分析步驟：

分析方法→存活分析→Kaplan-Meier 存活函數估計

→步驟一：資料匯入(選取個人資料檔 lung\_cancer\_study\_RT\_1)

→步驟二：參數設定(時間變數：SURVIVAL\_MONTH、事件變數：vital\_status)

→進階選項(選擇信賴區間方法：Log-log 與 Log)

→開始分析

(2) 檢定 " 僅 " 接受放療和 " 僅 " 接受化療病人間存活的差異，需先對資料進行篩選，篩選條件為

(CHEMO = Yes AND RT=No) OR (CHEMO = No AND RT=Yes)

，篩選後的資料共有 68 筆，篩選後的資料請另存新檔 lung\_CHEMO\_RT，存檔後的資料即可使用存活分析中的存活函數比較來檢定兩組的差異，在分析時可選用 CHEMO 或 RT 作為分組變數(因資料經過篩選，故兩變數皆可為分組變數)。

R-web 分析步驟：

分析方法→存活分析→存活函數比較

→步驟一：資料匯入(選取個人資料檔 lung\_CHEMO\_RT)

→步驟二：參數設定(時間變數：SURVIVAL\_MONTH、事件變數：vital\_status、分組變數：CHEMO 或 RT)

→進階選項(可加選其他檢定方法及存活函數圖比較兩組的差異)

→開始分析

(1)分析結果

- 分析方法：Kaplan-Meier 存活函數估計
- 資料名稱：lung\_cancaer\_study\_RT\_1
- 時間變數：SURVIVAL\_MONTHS
- 事件變數：vital\_status (設限指標：0)
- 顯著水準：0.05
- 信賴區間：Log-log
- 計算時間：0.275 秒

表 3 K-M 存活率估計及 Log-log(Log) 轉換的 95%信賴區間(RT=1)

時間 time	涉險人數 <sup>I</sup> no. at risk	事件人數 no. of event	K-M 存活率估計 K-M survival	標準差 std. err	95 % 信賴區間		95 % 信賴區間	
					Log-log		Log	
					下界 lower	上界 upper	下界 lower	上界 upper
2	70	1	0.9857	0.0142	0.9583	1	0.9029	0.998
3.25	69	1	0.9714	0.0199	0.9332	1	0.8906	0.9928
3.52	68	1	0.9571	0.0242	0.9109	1	0.873	0.986
5.65	67	1	0.9429	0.0277	0.89	0.9988	0.8549	0.9782
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
96.1	6	1	0.1688	0.0545	0.0897	0.3178	0.079	0.2875
114	4	1	0.1266	0.0548	0.0542	0.2958	0.0443	0.2539
115.8	3	1	0.0844	0.0502	0.0263	0.271	0.019	0.2139
132.5	1	1	0	NaN	NA	NA	NA	NA

表 4 百分位數估計值摘要(RT=1)

參數 parameters	估計 estimation
25 百分位數	77.6
50 百分位數	30.72
75 百分位數	16.8

圖 3 存活函數估計圖及 Log-log 轉換計算的%95 信賴區間(RT=1)

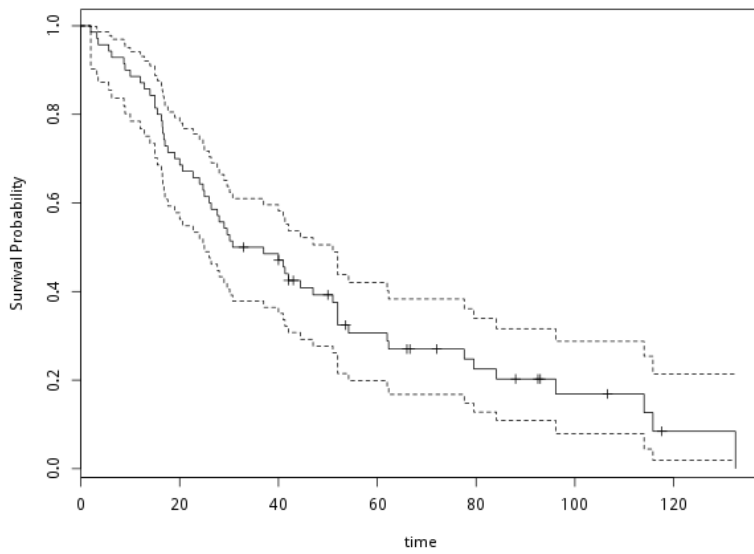
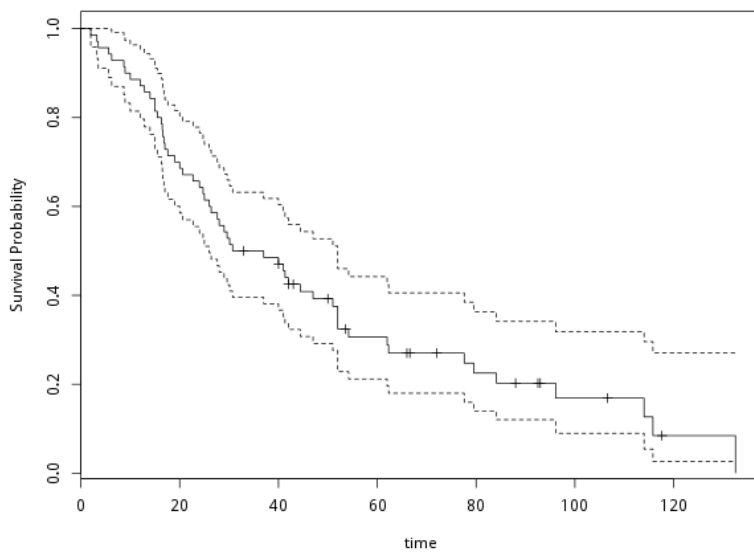


圖 4 存活函數估計圖及 Log 轉換計算的%95 信賴區間(RT=1)



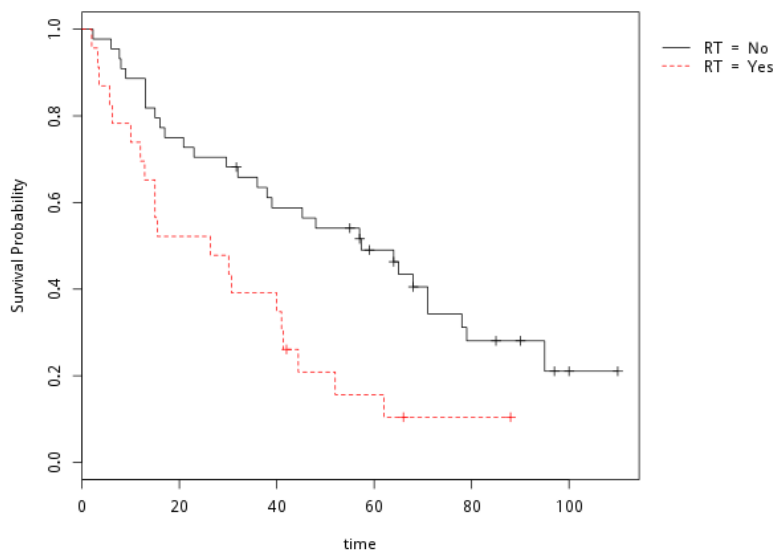


## (2)分析結果

- 分析方法：存活函數比較
- 資料名稱：lung\_CHEMO\_RT
- 時間變數：SURVIVAL\_MONTHS
- 事件變數：vital\_status (設限指標：0)
- 分組變數：RT (No, Yes)
- 計算時間：0.26 秒
- 存活函數比較：

虛無假設：各存活函數間無顯著差異			
檢定方法 method	檢定統計量 statistics	自由度 d.f.	p-值 p-value
Log-Rank 檢定	8.5999	1	0.0034
Gehan-Wilcoxon 檢定	8.1366	1	0.0043
Peto-Wilcoxon 檢定	8.149	1	0.0043
Harrington-Fleming 檢定( $\rho=0, \gamma=1$ )	5.6948	1	0.017

- 存活函數圖：



小結：

(1) 肺癌病人中接受化療 ( RT=1 ) 的存活曲線的 Kaplan-Meier 估計，信賴區間及圖分列於表 3 及圖 3、4，存活曲線的中位數及 $\frac{1}{4}$ 及 $\frac{3}{4}$ 分位數列於表 4 中。結果顯示與 CHEMO 的結論無異。

(2)比較檢定 " 僅 " 接受放療和 " 僅 " 接受化療病人間存活的差異，由四種檢定方法的結果顯示兩組的存活曲線有很大的差異，各檢定方法的 P 值界於 0.0034-0.017 之間，遠小於 0.05。由存活函數圖中可以看出僅接受 CHEMO 治療(RT=No)的存活率遠高於僅接受 RT 治療(RT=Yes)的存活率。

5. 將復發(relapse)到死亡時間當成“存活時間”，使用 COX 模型(I).做單一變數的分析，探討年齡(“年齡減年齡中位數”當做年齡變數)、性別、抽煙、化療、T-stage 是否為風險因子？風險比各為多少？95%信賴區間?(II).使用上述因子作多變量分析，並做出結論。(III).以性別做分層分析並給出結論。

答：

(I)此處為單變量分析，在使用 COX 迴歸模式將單一變量(如年齡，需先平減中位數)放入分析即可，重複此分析並在每次放入不同變量(性別、抽煙、化療、T-stage)，依據結果的 P 值來判斷是否為風險因子，R-web 分析結果亦會顯示風險比及 95%信賴區間。

R-web 分析步驟：

分析方法→存活分析→Cox 比例風險模式

→步驟一：資料匯入(選取個人資料檔 lung\_cancer\_study)

→步驟二：參數設定(時間變數：SURVIVAL\_MONTH、事件變數：vital\_status、共變數：年齡(或性別、抽煙、化療、T-stage))

→進階選項(可不設定)

→開始分析

此處僅以共變數 T-stage 為例，並解釋分析結果

(請注意資料中 T-stage 有兩處為空白，在上傳檔案前請先改成 NA)

## 分析結果

- 分析方法：Cox 比例風險模式
- 資料名稱：lung\_cancer\_study
- 時間變數：SURVIVAL\_MONTHS
- 解釋變數：T\_STAGE
- 事件變數：vital\_status (設限指標：0)
- 顯著水準：0.05
- 計算時間：0.309 秒
- 最終模式：

變數名稱 variable	係數估計值 coef. esti.	標準差 std. err.	z 檢定統計量 z statistic	p 值 p-value	估計值的指數 (風險比例) Exp(coef.) (Hazard Ratio)	Exp(coef.)的 95% 信賴區間	
						下界 lower	上界 upper
T_STAGE(T2 or T3)	0.1535	0.1794	0.8557	0.3922	1.1659	0.8203	1.6573
T_STAGE(T2, T3 or T4)	0.7536	0.1516	4.97	< 1e-04	2.1246	1.5784	2.8597

age_m	0.0272	0.0065	4.1888	< 1e-04	1.0275	1.0146	1.0407
-------	--------	--------	--------	---------	--------	--------	--------

GENDER(Male)	0.3287	0.1277	2.5749	0.01	1.3891	1.0817	1.784
--------------	--------	--------	--------	------	--------	--------	-------

SMOKING(Never smoked)	-0.4613	0.3155	-1.4625	0.1436	0.6304	0.3397	1.1699
SMOKING(Smoked in the past)	-0.2861	0.2505	-1.1425	0.2533	0.7512	0.4598	1.2272

CHEMO(Yes)	0.5295	0.1548	3.4211	6e-04	1.698	1.2537	2.2998
------------	--------	--------	--------	-------	-------	--------	--------

(II)此處使用 COX 迴歸模式分析多個共變量於模式中，將不同共變量(年齡、性別、抽煙、化療、T-stage)同時放入模式，依據結果的 P 值來判斷是否為風險因子，R-web 分析結果亦會顯示風險比及 95%信賴區間。

R-web 分析步驟：

分析方法→存活分析→Cox 比例風險模式

→步驟一：資料匯入(選取個人資料檔 lung\_cancer\_study)

→步驟二：參數設定(時間變數：SURVIVAL\_MONTH、事件變數：vital\_status、共變數：年齡、性別、抽煙、化療、T-stage)

→進階選項(可不設定)

→開始分析

## 分析結果

- 分析方法：Cox 比例風險模式
- 資料名稱：lung\_cancer\_study
- 時間變數：SURVIVAL\_MONTHS
- 解釋變數：age\_m, GENDER, SMOKING, CHEMO, T\_STAGE
- 事件變數：vital\_status (設限指標：0)
- 顯著水準：0.05
- 計算時間：0.307 秒
- 最終模式：

變數名稱 variable	係數估計值 coef. esti.	標準差 std. err.	z 檢定統計量 z statistic	p 值 p-value	估計值的 指數 (風險比例) Exp(coef.) (Hazard Ratio)	Exp(coef.)的 95% 信賴區間	
						下界 lower	上界 upper
age_m	0.0352	0.0083	4.2448	< 1e-04	1.0358	1.0191	1.0528
GENDER(Male)	0.0822	0.1643	0.5006	0.6167	1.0857	0.7868	1.4982
SMOKING(Never smoked)	-1.0705	0.3609	-2.9663	0.003	0.3428	0.169	0.6955
SMOKING(Smoked in the past)	-0.6056	0.2698	-2.2441	0.0248	0.5458	0.3216	0.9262
CHEMO(Yes)	0.7814	0.1741	4.4871	< 1e-04	2.1845	1.5528	3.0731
T_STAGE(T2 or T3)	0.0407	0.2232	0.1823	0.8553	1.0415	0.6725	1.6132
T_STAGE(T2, T3 or T4)	0.9116	0.1951	4.6719	< 1e-04	2.4883	1.6975	3.6474

(III)此處使用分層 COX 迴歸模式分析多個共變量於模式中，將不同共變量(年齡、抽煙、化療、T-stage)同時放入模式，此模式的分層變數為性別，依據結果的 P 值來判斷共變數是否為風險因子，R-web 分析結果亦會顯示風險比及 95%信賴區間。

R-web 分析步驟：

分析方法→存活分析→Cox 比例風險模式

→步驟一：資料匯入(選取個人資料檔 lung\_cancer\_study)

→步驟二：參數設定(時間變數：SURVIVAL\_MONTH、事件變數：vital\_status、共變數：年齡、抽煙、化療、T-stage)

→進階選項(分層變數：GENDER)

→開始分析



## 分析結果

- 分析方法：Cox 比例風險模式
- 資料名稱：lung\_cancer\_study
- 時間變數：SURVIVAL\_MONTHS
- 解釋變數：age\_m, SMOKING, CHEMO, T\_STAGE
- 事件變數：vital\_status (設限指標：0)
- 分層變數：GENDER (Female, Male)
- 顯著水準：0.05
- 計算時間：0.333 秒
- 最終模式：

變數名稱 variable	係數估計值 coef. esti.	標準差 std. err.	z 檢定統計量 z statistic	p 值 p-value	估計值的指數 (風險比例) Exp(coef.) (Hazard Ratio)	Exp(coef.)的 95% 信賴區間	
						下界 lower	上界 upper
age_m	0.0346	0.0083	4.1526	< 1e-04	1.0352	1.0184	1.0523
SMOKING(Never smoked)	-1.0266	0.3614	-2.8404	0.0045	0.3582	0.1764	0.7275
SMOKING(Smoked in the past)	-0.5653	0.2701	-2.0933	0.0363	0.5682	0.3347	0.9646
CHEMO(Yes)	0.7863	0.175	4.4936	< 1e-04	2.1952	1.5579	3.0932
T_STAGE(T2 or T3)	0.0354	0.2229	0.1586	0.8739	1.036	0.6693	1.6036
T_STAGE(T2, T3 or T4)	0.893	0.1948	4.5852	< 1e-04	2.4425	1.6674	3.5777

小結：

(I) 在單變量的 COX 迴歸分析中

(1)年齡(P 值 $<1e-04$ )顯著

(2)性別(P 值 0.01，相對於基準組 Female)顯著

(3)抽菸(Never smoked，P 值 0.1436 相對於基準組 Currently smoke)不顯著；抽菸(Smoked in the past，P 值 0.2533，相對於基準組 Currently smoke)不顯著。

(4)化療(P 值  $6e-04$  相對於基準組 CHEMO=No)顯著

(5)T\_STAGE(T1)為基準組，T\_STAGE(T2 or T3)與基準組比較 P 值(0.3922)不顯著，T\_STAGE(T2, T3 or T4) 與基準組比較 P 值( $< 1e-04$ )顯著。

(II)在多變量的 COX 迴歸分析中，共變量年齡、抽菸、化療及 T\_STAGE(T2, T3 or T4)為顯著，其 P 值皆遠低於 0.05，而性別及 T\_STAGE(T2 or T3)則為不顯著。

此結果與(I)中結果有差異，性別變數在單變量分析中為顯著，在多變量的分析中則是不顯著；而抽菸變數在單變量分析中為不顯著，但是多變量分析中則是顯著。由此處可看出性別與抽菸兩個變量可能存在干擾關係。

(III)在以性別為分層變數的多變量的 COX 迴歸分析中，共變量年齡、抽菸、化療及 T\_STAGE(T2, T3 or T4)為顯著，其 P 值皆遠低於 0.05，而性別及 T\_STAGE(T2 or T3)則為不顯著。此結果與(II)中一致，且係數估計變動不大，表示分層後與未分層的結論一致，不需要分層。